

July 2, 2001  
Date of Deposit

5 EL576314008US  
Express Mail Label

10 *Attorney Docket No. 21726/92526*

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
UTILITY APPLICATION

15 INVENTORS: Carol A. Westbrook and Ronald Hoffman

20 TITLE: **A cDNA DATABASE AND BIOCHIP FOR ANALYSIS OF  
HEMATOPOIETIC TISSUE**

ATTORNEY: Alice O. Martin  
BARNES & THORNBURG  
2600 Chase Plaza  
25 10 South LaSalle Street  
Chicago, Illinois 60603  
(312) 357-1313

09897798-070201

-1-

**A cDNA DATABASE AND BIOCHIP FOR ANALYSIS  
OF HEMATOPOIETIC TISSUE**

Inventors: Carol A. Westbrook and Ronald Hoffman

**5 BACKGROUND OF THE INVENTION**

This application claims priority from Serial No. 60/216829 filed July 7, 2000.

A unique database, a "transcriptosome" of a primate CD34+ cell, was compiled which is useful for the analysis of hematopoietic tissue. Research and clinical applications arise from analysis of bone marrow, peripheral blood or cord blood prior to gene therapy or transplantation of bone marrow, for example. Molecules with nucleotide sequences that are in the database may be placed in arrays on microchips for various applications.

Although the human genome has been sequenced, meaningful groupings and uses of the sequences are just beginning. Specific purpose databases (datasets) are not available for bone marrow and related tissues.

The concept of cDNA arrays has already been developed, and the technology is widely available. However, creation of databases by selecting genes according to a plan and/or specific uses or functions, to put on chips, is still an active area of research. An example is the "lymphoma chip" that was recently reported, which contained arrays of genes used for diagnosis of lymphoma (Alizadeh *et al.*, 2000).

To prepare an array so that it can be used for a specified purpose, some sort of support is generally needed. For example, cDNA chips are solid supports (usually glass slides or filter membranes) containing DNA fragments from a specific plurality of cDNAs, ESTs, or control molecules organized in 2-dimensional patterned arrays, which are used for hybridization to RNA or DNA probes. The chips are used, for example, to detect the presence, as well as the relative level of expression of each DNA of the array in a target sample. The technology of cDNA arrays and of signal quantitation is well-developed, but specific uses of the arrays, the nature of the DNA to be placed on the chips, and medical application of chips is still under investigation. Moreover, the term "chip" is becoming broad. "Microarray" means that a plurality of very small molecules are included.

-2-

**SUMMARY OF THE INVENTION**

The invention includes a database that is a set of nucleotide sequences for cDNA molecules including those for genes with known functions, in addition to genes with unknown functions, and ESTs (expressed sequence tags). The database is useful for the identification of genes relevant to hematopoiesis, and for the preparation of a microarray chip ("microchip" or "biochip") or other physical manifestation of an array that can be used to analyze hematopoietic tissue (bone marrow, peripheral blood, leukemia cells) for clinical applications such as bone marrow transplantation, and for research in human and other primate studies relating to hematopoiesis. The unique aspects of this invention include the method in which the genes were identified as significantly expressed in bone marrow, the preliminary and expanded gene list (the database), the concept of using the gene list as a stem cell or hematopoiesis-specific database, the concept of using the gene list for a cDNA chip, and the application of the cDNA chip for clinical and research purposes.

**BRIEF DESCRIPTION OF THE DRAWINGS**

FIG. 1 shows the correlation of gene expression between human and baboon CD34<sup>+</sup> cells. The normalized intensities of all the data points (25,920) from five releases of GeneFilters (GF200- GF204) hybridized to the baboon-derived CD34<sup>+</sup> probe were compared to those resulting from the human-derived CD34<sup>+</sup> probe by scatter analysis, using Microsoft Excel software.

FIG. 2 lists abundance categories of the common genes in human and baboon CD34<sup>+</sup> cells. A total of 15,407 cDNAs whose expression varies less than 3-fold between human and baboon CD34<sup>+</sup> RNAs were arbitrarily grouped into four relative expression categories, from low to very high abundance. The categories, based on the signal intensity of the human RNA relative filter background, are as follows: no expression (<3-fold), low abundance (3-fold to <10-fold), intermediate (10-fold to <25-fold), high (25-fold to <100-fold), and very high abundance (100-fold and higher).

FIG. 3 compares the expression level between human and baboon CD34<sup>+</sup> cells for genes selected from different abundance categories, by semi-quantitative RT-PCR. Five known genes representative of each of the abundance categories described in FIG. 2 were analyzed by RT-PCR using primers from the 3'-untranslated region of the gene. The PCR reactions were done with (+) or without (-) addition of reverse

-3-

transcriptase (RT) for the indicated cycle number (Cy). The genes tested are: *TM4SF4*, transmembrane 4 superfamily member 4; *PTK9*, protein tyrosine kinase 9; *CYP1B1*, cytochrome P450, subfamily 1 (dioxine-inducible), polypeptide 1 (glaucoma 3, primary infantile); *CSF3R*, colony stimulating factor 3 receptor; *B2M*,  $\beta_2$ -microglobulin. The intensity measured with GeneFilters was compared to that measured by RT-PCR.

FIG. 4 compares the expression level between human and baboon CD34<sup>+</sup> cells for apparent species-specific genes selected from Table 3. Representative analysis by semiquantitative RT-PCR for three transcripts from Table 3 with apparent species-specific expression as measured on GeneFilters, using primers designed from the 3'-untranslated region of the gene. The PCR reactions were done with (+) or without (-) addition of reverse transcriptase (RT) for the indicated cycle number (Cy). The intensity measured with Gene Filters (GF) is compared to that measured by RT-PCR, normalized to genomic DNA. Intensity ratio measurement are shown as positive when expression in humans is higher than baboons, and negative when the reverse is true.

#### DESCRIPTION OF THE INVENTION

The invention relates a database ("transcriptosome") of a primate CD34<sup>+</sup> cell that includes sequences selected by methods of the present invention.

Because the database contains many unknown and uncharacterized genes, an important use of the invention is to discover new genes that are relevant to hematopoiesis and stem cell growth. The database also has value because it could be mined for specific gene discovery, for example to find new genes that are surface markers (e.g. for flow cytometry), growth factors, or receptors for growth factors that regulate stem cell growth. The database itself may have commercial use in its entirety for the preparation of chips, which could be used to diagnose or analyze hematopoietic cancers, and to evaluate normal bone marrow or stem cells prior to transplantation.

More particularly, the invention relates to a database that is a dataset which specifies the majority of genes expressed at moderate levels or higher in human hematopoietic tissue, as represented by CD34<sup>+</sup> cells from bone marrow, and their approximate rank order by level of expression. The genes in this database refer to

-4-

partial sequences that are available in the Human Genome databases, and thus can be analyzed directly by reference to their unique ID numbers. The database has value because it can be mined to identify abundant mRNAs coding for proteins of interest in many categories with therapeutic, research, and diagnostic applications. The gene list, or a subset thereof, is useful to prepare a cDNA chip with applications to hematopoiesis.

Alternatively, the gene list can be mined without preparing a chip from it. The preparation of a chip is one aspect of the invention and use of the database.

An aspect of the invention is a standard size cDNA chip (5,000 to 10,000 elements) constructed to contain genes expressed in human bone marrow, specifically those that are expressed in the CD34+ fraction, the fraction which contains the undifferentiated cells that give rise to stem cells and which contains transplantable elements. The cDNA composition of a chip made in this fashion is representative of genes that are expressed at moderate to high levels by human bone marrow cells in their native stage (natural, *in vivo*), and those genes whose expression might change with physiologic or pharmacologic manipulation, as well as those genes used as internal controls. However, other compositions of cDNA molecules are within the scope of the invention.

The invention also relates the composition of a chip, that is, the selection of DNA molecules to array (position on the support in accord with a plan, or strategy) on the chip, which is based on the results of a novel experimental method. The invention also specifies some of the uses of the chip, which include analysis of human bone marrow, peripheral blood or cord blood prior to transplantation to determine if the transplanted tissue will engraft; analysis of human bone marrow, peripheral blood or cord blood after it has been treated with approved or experimental manipulations (*e.g.* growth factors, purging, gene therapy, and the like) prior to transplantation, to determine if the transplantation will engraft, or to determine the effects of treatment; research in human bone marrow transplantation and *ex vivo* cellular expansion; discovery of new genes related to human hematopoiesis or stem cell growth; similar research in non-human primate system, with the aim of applying the research results to human systems.

-5-

5 A cDNA chip called, for example, the "Stem Cell Chip" is useful as a substrate for hybridization of RNA derived from human clinical or research samples, including hematopoietic stem cells obtained from sources such as bone marrow, peripheral blood, or cord blood; or from similar samples obtained from primate bone marrow for research purposes. The term "the chip" used hereinafter includes a plurality of chips either of similar or different compositions.

10 RNA is used to prepare a probe using standard methods (reverse-transcription, labeling by fluorescent or radioactive nucleotides), and the RNA is hybridized to the Stem Cell Chip. Hybridization occurs between homologous sequences - the degree of homology required for hybridization depends on the conditions under which the hybridization takes place, *e.g.*, temperature, pH. Hybridization to each cDNA molecule on the array is detected and quantitated. The pattern and the relative intensity of hybridization of the probes with each cDNA on the array is expected to vary with the population tested. Individual hybridization patterns and intensity levels  
15 define "clusters" of gene expression that are used to define physiologic conditions. For example, the chip may be applied to analyze a bone marrow that was treated with gene therapy, to determine if the marrow is likely to engraft for transplantation. The expression of genes on the chip would be compared to that level of expression needed for a successful graft. Another novel use of the chip is the study of experimental  
20 methods applied to non-human primates, particularly baboons. Because the chip is expected to be similarly representative of both human and baboon marrow, the use of this chip to analyze baboon marrow (stem cells or cord blood) makes it possible to directly apply the animal results to human systems. Because the chip may contain many uncharacterized gene fragments in the form of ESTs, an important use is in the  
25 discovery of new genes that are relevant to hematopoiesis and stem cell growth. Their relevancy is based on their inclusion on the gene list, and also by experimental uses of the chip such as to determine results of treatment, or comparisons of populations.

**Highly-abundant genes in the transcriptosome of human and baboon CD34 antigen-positive bone marrow cells**

30

Non-human primates are useful large animal model systems for the *in vivo* study of hematopoietic stem cell biology. To ascertain and analyze the degree of similarity of the hematopoietic systems between humans and baboons, and to explore

09897798.070201

-6-

the relevance of such studies in non-human primates to humans, the global gene expression profiles of bone marrow CD34<sup>+</sup> cells isolated from these two species were compared. Human cDNA filter arrays containing 25,920 human cDNAs were surveyed. The expression pattern and relative gene abundance of the two RNA sources was similar, with a correlation coefficient of 0.87. A total of 15,970 of these cDNAs were expressed in human CD34<sup>+</sup> cells, of which the majority (96%) varied less than 3-fold in their relative level of expression between human and baboon. RT-PCR analysis of selected genes confirmed that expression was comparable between the two species. No species-restricted transcripts have been identified, further reinforcing the high degree of similarity between the two populations. A subset of 1554 cDNAs which are expressed at levels 100-fold and greater than background is described, which includes 959 ESTs and uncharacterized cDNAs, and 595 named genes, including many that are clearly involved in hematopoiesis. The cDNAs reported here represent a selection of some of the most highly-abundant genes in hematopoietic cells, and provide a starting point to develop a profile of the transcriptome of CD34<sup>+</sup> cells.

Non-human primates are important experimental models for hematopoietic stem cell transplantation and biology, because the behavior of hematopoietic stem and progenitor cells in primates closely resembles that in man (Andrews *et al.*, 1992; Brandt *et al.*, 1999; Goodell *et al.*, 1997). The use of non-human primates permits a degree of experimental freedom to perturb hematopoiesis not possible in man, which might end in a genetic analysis of hematopoiesis, not only under steady-state conditions, but also under conditions of stress. The baboon (*Papio anubis*) is particularly useful in this regard because it is closely related to humans, and shows cross-reactivity with many of the reagents used to study human hematopoiesis. Recent studies have initiated a description of the overall pattern of gene expression in murine bone marrow stem cells (Nachtman *et al.*, 2000; Phillips *et al.*, 2000), but by contrast, relatively little is known of the expression patterns of human bone marrow hematopoietic stem cells or the baboon marrow stem and progenitor cells. To study baboon hematopoiesis, and facilitate extrapolation into human systems, the expression profiles of human tissue for each species were compared. Human and baboon bone marrow cells which were positive for the CD34 antigen (CD34<sup>+</sup> cells) were used for

-7-

these comparisons, because they represent a marrow fraction enriched for both primitive hematopoietic stem and progenitor cells (Link *et al.*, 1996; Pierelli *et al.*, 2000; Ueda *et al.*, 2000).

- Human cDNA filter arrays were used to establish the expression profiles for both species, because there is no comparable product available for baboon cDNA analysis, and a high nucleotide sequence homology between these two species was expected (Liao *et al.*, 1998; Trezise *et al.*, 1989). The cDNA filter arrays used (GeneFilters™) contained 25,920 cDNAs from the UniGene dataset (<http://www.ncbi.nlm.nih.gov/UniGene/index.html>), including both known genes and uncharacterized ESTs, permitting the survey of one-fourth to one half of the estimated 50,000-100,000 genes in the genome. The transcriptome of CD34<sup>+</sup> cells, is disclosed herein, demonstrating very comparable gene expression patterns in CD34<sup>+</sup> cells in these two species, and validating the utility of human cDNA arrays for baboon studies.
- SELECTION OF THE GENE LIST (database): The gene list (database) of this invention was defined using a unique approach combining filter array methodology with cross-species hybridization to identify conserved sequences. Normal human bone marrow from an anonymous donor was fractionated into CD34<sup>+</sup> cells by standard methods (using anti-CD34<sup>+</sup> antibody to bind and separate out cells). RNA was prepared from the CD34<sup>+</sup> cells so obtained, and then used to prepare a hybridization probe by radioactive labeling; the probe was hybridized to a commercially-available cDNA filter array (GeneFilters, release 200 - 204, purchased from Research Genetics, Huntsville, AL), which contained in total 25,900 cDNAs and ESTs from the UniGene set. The 25,900 genes surveyed represent 1/3 to 1/2 of the estimated 50,000 to 75,000 genes in the human genome. After hybridization of the arrays to the human CD34<sup>+</sup> RNA probe, similar probes were prepared from normal baboon marrow cells that had been similarly purified for CD34<sup>+</sup> cells. Comparison of the hybridization profiles of the human and baboon marrow made it possible to determine that both had similar expression patterns for the majority of genes. The use of a cross-species hybridization (human and baboon) ensured the selection of genes that were conserved between both species. Thus, the selected genes which are present in both RNAs are expected to be more representative of the tissue, ie. CD34<sup>+</sup> cells, than of the individual species. The

-8-

- correlation of human and baboon marrow varied from 88% to 98%, depending on the filter analyzed, with an average correlation of 94%. (To put these figures in perspective, a correlation coefficient of 0.42 was measured when comparing CDE34+ expression on GeneFilters to that obtained for the hematopoietic cell line U937 and a correlation coefficient of 0.57 when comparing human CD34+ cells to HT29 colon cancer cell line).

- A set of approximately 9,500 genes was selected using two criteria: all of those expressed at similar levels in both human and baboon (which was defined as a level of expression that varied 3-fold or less between the species) and whose expression in the human was 7-fold or greater than the background level that was measured in the individual GeneFilter experiment (which was arbitrarily assigned to indicate expression at a moderate to high level). A cut-off level of intensity of 3-fold over background is generally taken to indicate expression that is greater than zero, and can be reliably detected and quantitatively measured for the human-based probes.
- Using this cut-off of 3-fold, the human CD34+ cells displayed approximately 15,970 or 62% of the 25,920 cDNAs present on these filters. The level of 7-fold over background was thus arbitrarily selected as a cut-off for this gene list, recognizing that all of these genes are certain to be actually expressed in the cells, and to provide a dataset that was limited in size to <10,000 genes, and contained those that are expressed at moderate to high levels; a more complete dataset would include the entire 15,970 genes; by extrapolation, this may represent half to third of all of the genes in the CD34+ cells. For some applications, different cut-off levels could be utilized--a higher cut-off would result in fewer genes but they would be a high level, and a lower cut-off would be more inclusive of the entire expression profile of the cell.

- Genes from this database were then ranked from highest to lowest level of expression, as determined from their measured intensity in human CD34+ RNA. The rank order is only approximate, because the filters cannot provide the absolute level of expression, and there is experimental error in taking the measurements, but confirmatory experiments for randomly-selected genes have shown a fairly good correlation with rank order and expression measured by other methods. Additions, or corrections to the list may be made within the scope of the invention, but the

-9-

underlying concept and the majority of the listed genes are as indicated herein. The complete gene list is appended as Appendix A and is available through a web site <http://westsun.hema.uic.edu/html/expression.html> which will be available to the public upon filing the present patent application. Table 2 shows selective highly-abundant EST's and partially characterized cDNAs in human an baboon CD34+ cells.

The gene filters which were used to identify the genes are commercially available from Research Genetics, but any filter array might have been used. The genes themselves are selected from databases that are in the public domain (UniGene dataset, <http://www.ncbi.nlm.nih.gov/UniGene/index.html>) as part of the Human Genome Program. The invention is to compile a specialized database using the criteria herein for applications involving hematopoietics.

The genes defined in this invention are represented as UniGene cluster numbers. UniGene (<http://www.ncbi.nlm.nih.gov/UniGene/index.html>) is a product of the Human Genome Program, maintained by the National Center for Biotechnology Research. UniGene contains over 40,000 entries, each of which represents a unique gene based on a composite of sequences of individual clones from cDNA libraries. The cDNA clones represented in UniGene are available for purchase from a number of repositories, including TIGR (The Institute For Genome Research, <http://www.tigr.org/tdb/tdb.html>). The dataset and representative clones are publicly available to any investigators, but the clones specified by this invention, and their association as a group with bone marrow and related cell types, and their expression levels, are not publicly available data.

Furthermore, there is currently no commercially available cDNA chip that has genes representative of human bone marrow stem cells and related cell types, nor is there such an extensive database which describes the constitution of genes expressed in human bone marrow. Furthermore, until the present invention, it was not possible to directly translate research results from experimental primate studies (baboon) to humans.

Table 1 shows some of the most abundant cDNAs commonly expressed in human and baboon CD34+ cells. This table displays the first 200 genes from the total genes in Appendix A, or the top 2% (by expression level). Table 1 is derived from the Appendix, that contains the entire gene set, that is those that are >7-times over

-10-

background in human and less than 3-fold different between species. The column headings, from left to right are:

1. **Rank order** (based on human expression).
- 5 2. **CLUSTER ID** (refers to the human Unique Gene number, or UniGene number, part of the Human Genome Program.  
<http://www.ncbi.nlm.nih.gov/UniGene/index.html>)
3. **GENBANK** the GenBank number of the clone from the UniGene cluster which was placed on GeneFilters and which hybridized to the probe
- 10 4. **Human expression level** (measured experimentally, as normalized intensity).
5. **Baboon expression level** (measured experimentally, as normalized intensity).
6. **Relative expression level**, expressed as a ratio of human to baboon, from experimental data.
- 15 7. **Title-** name of gene or EST, extracted by Pathways software (Software from Research Genetics used to interpret the GeneFilters Result) from the UniGene databases.
8. **Official gene name**, if known.

Note that columns #2,3, 7 and 8 may be updated as the UniGene databases are updated, but they still refer to the same gene.

## EXAMPLES

### Example 1: Use of the Hematopoietic Database of the Present Invention to Expand a Stem Cell Graft *Ex Vivo*

25 A use of the database is to determine whether a stem cell graft has the same level of gene expression as the host, or desired stem cells, in particular for genes known to be related to the success of expansion of a stem cell graft *ex vivo*. To do this, the pattern of gene expression in the host stem cells for genes in the database of the present invention must be analyzed. A comparison is then made of the level of expression of the same genes, in the graft. An embodiment of the invention is to compare expression levels of genes of a subset of genes either highly expressed in stem cells, or known to be predictive of stem cell graft expansion success.

09897798.070201

-11-

**Example 2: Use of the Hematopoietic Database of the Present Invention to Determine Whether or Not Genetic Modification Altered the Molecular Signature of Tissue**

5           Gene therapy is used to alter or replace defective genes or to enhance the expression of specific genes.

          To determine whether genetic modifications did or did not alter the molecular signature of tissue used in gene therapy, expression levels of genes in the database of the present invention are compared before and after the modifications are made.

10   **MATERIALS AND METHODS**

**I. Collection and Selection of CD34<sup>+</sup> marrow cells**

          Healthy adult baboons (*Papio anubis*) weighing 9-10 kg were used. The animals were housed under conditions approved by the Association for the Assessment and Accreditation of Laboratory Animal Care. Bone marrow aspirates  
15   were obtained from the humeri and iliac crest of adult baboons under ketamine and xylazine (1 mg/kg) anesthesia under guidelines established by the Animal Care Committee of the University of Illinois at Chicago. Human bone marrow aspirates from the iliac crest were obtained from normal human adult donors after informed consent was obtained, as approved by the Institutional Review Board of the University  
20   of Illinois at Chicago. Marrow mononuclear cells were isolated from the marrow as previously described (Brandt *et al.*, 1999). Briefly, the marrow was heparinized; diluted 1:15 in phosphate-buffered saline (PBS); and fractionated over 60% Percoll (Pharmacia LKB, Uppsala, Sweden) by centrifugation at 500 g for 30 minutes at 20°C. The interphase mononuclear cells were resuspended in PBS containing 0.2% bovine  
25   serum albumin and human immune globulin (Sigma Chemical Co, St. Louis, MO) and labeled with the biotin conjugated mouse anti-human CD34<sup>+</sup> antibodies MoAb 12-8 (Andrews *et al.*, 1986) for baboon, and QBAND/10 (Brandt *et al.*, 1998) for human cells, washed, and relabeled with streptavidin conjugated rat anti-mouse antibody-containing iron microbeads (Miltenyi Biotec, Auburn, CA). The CD34<sup>+</sup>  
30   cells were then selected by passing the CD34<sup>+</sup> cell-antibody-iron bead complex through a magnetic column. The purity of the CD34<sup>+</sup> fraction was estimated by flow cytometry using a fluorescein isothiocyanate (FITC)-conjugated anti-human CD34<sup>+</sup> antibody K6.1 (Brandt *et al.*, 1999) for baboon cells and MoAb HPCA-2 for human cells.

09697758.070201

-12-

## II. RNA and DNA preparation

Total RNA was extracted from  $1-5 \times 10^6$  human and baboon CD34<sup>+</sup> cells using an Ultraspec RNA Isolation kit (Biotecx Laboratories, Inc, Houston, TX) according to the manufacturer's protocol. The quantity of total RNA was determined by  $A_{260}$

- 5 absorbance, and quality was verified by analysis on 1% agarose gels using standard techniques. Genomic DNA was prepared from the HL60 human cell line (American Type Culture Collection) and baboon peripheral blood cells using Trizol reagent (Life Technologies) according to the manufacturer's specification.

- 10 Uniformly-labeled cDNA probes were prepared from 3 mg of total RNA by priming with 2 mg of oligo-dT, followed by elongation with 1.5 units of Superscript II reverse transcriptase (Life Technologies, Grand Island, NY) in presence of 100 mCi of  $^{32}\text{P}$  dCTP (Amersham Pharmacia Biotech, Piscataway, NJ). The labeled probe was purified from unincorporated nucleotides and other small molecules with ProbeQuant G-50 (Amersham Pharmacia Biotech).

## 15 III. Hybridization of cDNA probes to GeneFilters

- Five releases (GF200-204) of human GeneFilters (Research Genetics, Huntsville, AL) were pre-hybridized for 2 hours at 42°C in MicroHyb solution (Research Genetics), with the addition of 1  $\mu\text{g/ml}$  each of polyA (Research Genetics) and human Cot1 DNA (Life Technologies, Grand Island, NY). The blots were then
- 20 hybridized overnight in the same MicroHyb solution with the addition of  $2 \times 10^6\text{cpm/ml}$  of heat denatured probe. The blots were washed twice at 50°C with 2X SSC, 1% SDS for 20 minutes and once at room temperature in 0.5X SSC, 1% SDS with gentle agitation for 15 minutes, prior to imaging. For re-use of membranes, the filters were stripped in 0.5% SDS for 1 hour at room temperature with gentle agitation
- 25 as recommended by the manufacturer, and was re-exposed to confirm complete stripping.

## IV. Exposure, Imaging, and Analysis of Filter Membranes

- The hybridized filters were imaged using a phosphor imaging screen (Molecular Dynamics, Sunnyvale, CA), exposed for three to four days, imaged using a
- 30 Storm phosphor imaging system (Molecular Dynamics) at 50-micron resolution, and analyzed using PathwaysII from Research Genetics following the manufacturer's guidelines. Using this program, individual cDNA spots were identified and fit to a

09897798.070201

-13-

grid, and their intensity measurements were recorded as raw intensities. The background for a particular experiment, provided as a reference, was calculated by averaging the measured intensities between the two grids of the filter. This background information was used to assign levels of expression of the genes. Data from poor hybridizations, such as those which had unacceptably high background or non-uniform control spots intensities across the membrane, was not considered for further analysis and discarded. To compare expression of a cDNA spot between two probes that were sequentially hybridized to the same filter, the intensities were normalized using the algorithm provided by the PathwaysII software, using either control spots or all data points as reference. The data were exported as Excel files for further analysis. Since PathwaysII utilizes an older, somewhat outdated version of UniGene (build versions 18, 19 ,39, and 42) and substantial changes have been made in the UniGene database since then, the cDNAs list was updated using UniGene build version 118 as reference (current as of April, 2000). To accomplish this, both the UniGene and GeneFilter dataset were reformatted to Microsoft Access database. The GenBank accession numbers of the GeneFilter dataset were then matched against the UniGene database to update the cluster ID, gene name, and gene description.

#### V. PCR Analysis

For reverse-transcriptase PCR (RT-PCR), first strand cDNA was generated from approximately 1 mg of RNA that had been DNase-treated with RNase free DNase I (Life Technologies, Grand Island, NY). The RNA was then used to make first strand cDNA in a 20 ml reaction volume with (+RT) or without (-RT) reverse transcriptase using Superscript II Reverse Transcriptase kit from Life Technologies according to the manufacturer's recommended protocol followed by RNase H treatment. If not stated otherwise, 1/20th volume of the +/- RT reaction mix was used for the PCR reaction in presence of 1X PCR buffer (Perkin Elmer Cetus (PE)), 1.5mM MgCl<sub>2</sub>, 200mM dNTPs, 1 mM each of forward and reverse primers, and 1U of Amplitaq polymerase (PE ) in a 20 ml reaction volume using the following cycles; initial denaturation at 95°C for 5 min. followed by each cycle at 95°C for 30sec., annealing at 58°C /65°C depending on the primer pair for 30sec., amplification at 72°C for 30sec., the final amplification was for 5 min at 72°C. PCR analysis of

## VI. Comparison of expression levels by semi-quantitative RT-PCR

To compare the expression of individual genes, RT-PCR was performed using primer pairs designed based on the sequence of the cDNA clones that was included on the GeneFilter. The PCR was done from 25 to 40 cycles with increments of 5-cycles, except for  $\beta_2$ -microglobulin, which was done at 18, 22, 25, and 30 cycles. The PCR reaction products were analyzed on a 3% agarose gel stained with ethidium bromide, and the amount of DNA was quantitated as band intensities using GelDoc software from BioRAD (Hercules, CA). The level of expression of each gene was normalized against the level of  $\beta_2$ -microglobulin expression between these two species. The relative expression between human and baboon cDNA was estimated by measuring the ratio of intensity of DNA product, comparing only those measurements which fell within the linear range of PCR amplification cycles; multiple determinations, when performed, were averaged. The sequences of Forward (F) and Reverse (R) primers are: Transmembrane 4 superfamily member 4 (TM4SF4), F-AAGCGATTTCGATGTTACCTC, R-GAGGCTCTCGGCACTTGTTCC; Protein tyrosine kinase 9 (PTK9), F-GATTCCTTTGTTTTACCCCTGTTGGAG, R-TTGCTGC ATACAACATTTTTGAC ; Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1 (glaucoma 3, primary infantile) (CYP1B1), F-GTAATGGTGTCCAGTATAA GTAATGAG-3', R-TCATGAATGCTTTTAGTGTGTGC-3'; Colony stimulating factor 3 receptor (granulocyte) (CSF3R), F-CTGAAGTTATAGGAAACAAGC ACAAAGGC, R-GCCC ATGACTAAAACTACCCAGC; Beta-2-microglobulin (B2M), F-CCTGAATTGCTA TGTGTCTGGG, R- TGATGCTGCTTACATGTCTCGA. R82595, F : GCTCGTAGCAACATTTTCGTAATAGCC, R : GGACCCATCGTGGTT ACCGTG; AA676327, F-ATATTTTCGGTAACTTTGTACCCCTAAG, R: CAGGGGCAA TTTTGAGGTATG; R85439, F: GGCAGGGCTCTAAATGGAAGTAGTTG, R: CTCAG AAGTGTTTTGTAGCAAGGCTGC, AA487912, F: AAACAGTGACTTATCCCGCTAC CC, R: GGGTGGGTTTACTCTTAGAATCGC; N25920. F: CAGATGGAGGGTTTATG

-15-

AGTGAGGCTGG, R: GCTTGTTCTTTGGGGATTGTGGTGC; R05886, F:  
TAGGCG TGAGAAGCATATAGAGGC, R: AGTGAATAAGCAAGAAATCAGGGTG; N74363, F:  
ACAAAGGGCTGTTTACTGAGAGACCTGAGC, R:  
GGCATAACTCACACCCATT TGTTTACCTGC; N55359, F:  
5 GGCAGAATCTACTGGGCATCTTGTAATC,  
R: AGTTTGGTGGTCCAGGGAAGGTAC.

**VII. Correlation of gene expression between human and baboon CD34<sup>+</sup> cells**

CD34<sup>+</sup> cell populations were isolated from bone marrow aspirates by immunomagnetic cell sorting using antibodies that represent the best selection of  
10 undifferentiated and multi-potent marrow cells in human and baboon marrow. The human marrow cell population was 90% pure, as determined by FACS analysis with anti-human CD34<sup>+</sup> antibody. Using the same method, the baboon CD34<sup>+</sup> cells measured 77% purity. This measurement in baboon cells is an underestimate of the true degree of purity due to the relative non-specificity of the anti-human CD34<sup>+</sup>  
15 antibody K6.1 (used for quantitation by flow cytometry) with baboon cells, resulting in a weaker fluorescence signal and lower estimates of purity than can be measured in comparable human cells, but it is within the range that we normally observe with this method.

Radioactively-labeled RNA-based probes prepared from each cellular  
20 population were hybridized to five nylon filter membrane arrays (GeneFilters releases 200-204, containing a total of 25,920 cDNAs) and phosphoimaged, and the resultant image was analyzed to determine the relative hybridization signal intensity for each cDNA with each probe. Each cDNA on the array is derived from a single clone from the IMAGE consortium (<http://image.llnl.gov>) representing the 3'-end of a unique  
25 UniGene cluster. All data were obtained by sequential hybridization to a single filter set, in order to provide the most accurate comparisons between probes and avoid variability in cDNA spotting. Duplicate experiments were performed when possible, but were limited by the lifetime of the filters, which in general could be successfully re-hybridized no more than 3 to 4 times. It was not possible to use pooled baboon  
30 marrow donors because of the limited availability of animals, and thus pooled human donors were not used either, recognizing that the methods of the present invention are not sensitive enough to detect small differences between individual donors.

09897798-070201

-16-

Normalized signal intensities for individual cDNA spots from these hybridizations were compared by scatter analysis, and revealed that the gene expression patterns in human and baboon cells were very similar, with an overall correlation of 0.87. The composite data for all hybridizations is summarized on a scatter plot (FIG. 1). The measured raw intensity of the hybridization signal relative to the filter background is used as an indicator of the relative abundance of the cDNA. For these experiments, a cut-off level of raw intensity (non-normalized) of 3-fold over background was used to indicate that a gene is definitively expressed in human cells. By this criteria, human CD34<sup>+</sup> cells displayed positive expression for approximately 15,970 (62%) of the 25,920 cDNAs present on these filters. This gene list excludes many housekeeping genes, which are measured on the GeneFilters as hybridization controls but are not included for normalization by Pathways II software. (For information on all the spotted cDNA for each filter including the housekeeping genes, refer to the Research Genetics's ftp website,

15 <ftp://ftp.resgen.com/pub/genefilters/>.

The baboon-derived probes showed a consistently higher hybridization background, approximately three-fold higher, than the human-derived probes, so it was not possible to apply the same cut-off level for this species (baboon). However, 13,447 cDNAs (84%) gave a signal with the baboon probe that varied less than 2-fold from the human level of expression, while almost all of the genes (15,407 or 96.5%) were expressed within 3-fold of each other. Much of the measured differences in expression level is likely to be due to experimental variation; about 3% of cDNAs will vary more than 3-fold upon repeat hybridization with these probes. Other measured differences between the human and baboon RNAs probably reflect true differences in expression, but in either case, the variation is not great. Thus human and baboon CD34<sup>+</sup> cells express virtually the same spectrum of genes, with similar though not identical levels of expression.

#### **VIII. cDNAs highly expressed in both human and baboon**

The 15,407 cDNAs that are commonly expressed in human and baboon CD34<sup>+</sup> cells were arbitrarily placed into several groups (FIG. 2) based on their spot intensities relative to background in the human data set: very high abundance (100-fold and over), 1,619 cDNAs; high abundance (25-fold to <100-fold), 2,376 cDNAs;

-17-

intermediate abundance (10-fold to <25-fold), 2,976 cDNAs; low abundance (3-fold to <10-fold), 8,436 cDNAs.

The very highly-abundant genes identified by Pathways II analysis were then updated to the most current UniGene release (version 118, April 2000), and examined in detail. A total of 1,554 UniGene clusters remained after updating. This list included 595 named genes, and 959 ESTs and uncharacterized cDNAs. This list of highly-abundant genes and ESTs is available as an appendix to the online version of this article, and is also available on our hematopoietic stem cell website (<http://westsun.hema.uic.edu/html/expression.html>). The named genes represent a wide variety of functional categories such as growth factors and cytokines, receptors and cell surface molecules, intracellular signalling molecules, cell cycle proteins etc. A sample of these genes, sorted by functional category, are given in Table 1. Note that this list includes many of the genes (typed in bold) that would be expected to be present in CD34<sup>+</sup> cells, such as receptors for IL3 and colony stimulating factor 3. Interestingly, many expected hematopoietic genes are not in this category, as their level of expression is relatively low; for example, the CD34 antigen is expressed at a relatively low level, only 6-fold above background (for human).

A large fraction, over 61% of these highly-expressed cDNAs, are ESTs and uncharacterized cDNAs. Although many of these genes are uncharacterized, the UniGene database provides some information about their similarity to known proteins. Furthermore, many of the named genes represent full length cDNAs that have not been fully studied or are only partially characterized, though some function is suggested by homology to known proteins. A partial list of some of these interesting ESTs and partially characterized named genes are given in Table 2. Further characterization of the ESTs in this database represents a potential wealth of new information about the CD34<sup>+</sup> transcriptome.

Several known genes from each abundance category were selected to verify their relative level of expression in both species by semi-quantitative RT-PCR. Representative examples are shown in FIG. 3. Each gene tested was found to be expressed at comparable levels in both species, although the abundance category was not always accurate, especially in the lower abundance genes. For example, PTK9 is expressed at a level 5-fold above background in human cells, but its signal appears

-18-

stronger than CYPB1, measured at 20-fold above background. The measurement of the absolute level of expression of a cDNA using filter hybridization is related to many factors, including the amount of DNA placed on the filter (which cannot be accurately controlled), and the efficiency of hybridization. Thus, the assignment of a gene to a relative abundance category can only be regarded as approximate, and may require additional confirmation.

#### **IX. Species-specific transcripts**

Although there were a number of cDNAs which did not appear to be highly-correlated (that is, their expression varied more than 3-fold between species), there were a few genes whose measured intensity suggested that they were preferentially expressed in only one species. To identify these genes, the GeneFilters dataset was searched for cDNAs which were unexpressed in one species (defined as a raw intensity of less than 3-fold background), and were clearly expressed in the other species (> 3-fold background) with a normalized intensity ratio of >3 fold between species. There were only 14 cDNAs which fit this criteria, 6 baboon and 8 human, which includes 6 known genes and 8 ESTs. PCR primer pairs for all 14 cDNAs were designed to match the sequence of the human clones which were present on the filter membrane; the pairs were tested for their ability to amplify both genomic DNA and reverse-transcribed RNA from both species. Six primer pairs (4 human and 2 baboon) were successfully validated on both species in this manner, and these were further analyzed by semi-quantitative RT-PCR, using an additional normalization factor for PCR efficiency on genomic DNA from both species. The ratio of expression for each gene, as measured by semi-quantitative RT-PCR, is compared to that measured on GeneFilters, is summarized in Table 3, and representative examples are shown in FIG. 4. The use of normalization factors, one as a control for PCR efficiency of human-specific primers against baboon, and another for RT-reaction, adds complexity and probably some inaccuracy in quantitative comparison of gene expression between the two species, so the measured levels can only be regarded as estimates. Nonetheless, most of the genes, except for two designated by Unigene Cluster ID Hs.1817 and Hs.215595, showed little if any differential between the two species and fall within 3-fold of each other, well within the arbitrary cut-off that was set for Table 1. Only Hs.1817 and Hs.215595 were confirmed to be expressed at somewhat higher levels in

-19-

human than baboon (3.6-fold and 5.4-fold, respectively), although the differences were small and not as great as was measured on the filters. The results showing differential expression of Hs.1817 are included in FIG. 4. Thus, none of the 6 genes tested showed expression restricted to one species, though some appear to be differentially expressed. This result suggests that the experimental variation in the GeneFilter hybridization system is greater than the actual variation between the two species. Additional work will be required to determine if there are any bonafide species-specific genes within either species.

By its ability to simultaneously detect and quantitate the expression level of thousands of genes at one time, cDNA array technology is greatly improving our understanding of the complex patterns of gene expression in eukaryotic cells. In the present invention this technology is used to profile the gene expression patterns of CD34<sup>+</sup> marrow cells in human and baboon cell populations. Baboon-derived probes are suitable for use on human cDNA arrays with some limitations.

Expression studies on cDNA arrays require a fairly large number of cells to isolate an appropriate amount of RNA for probe preparation. Because of this constraint, it was necessary to purify the CD34<sup>+</sup> cells by immunomagnetic columns rather than FACS, which would require prolonged sorting. The stress imposed by the prolonged sorting time required to prepare this number of cells can dramatically reduce cell viability and yield of CD34<sup>+</sup> cells, and may alter their gene expression profile. Because of the weak cross-reactivity of anti-human CD34<sup>+</sup> antibody against baboon CD34<sup>+</sup> antigen, it is difficult to accurately determine the level of purity of baboon CD34<sup>+</sup> cell population. Thus, the purity of baboon CD34<sup>+</sup> may be an under-representation. At any rate, in spite of the heterogeneity of the cell populations examined and the limited number of subjects studied, we determined that bone marrow cells derived from the two closely-related species have similar patterns of gene expression. Although many molecular similarities were expected between human and baboon CD34<sup>+</sup> cells, the results suggest that the transcriptomes are nearly identical, supporting experimental studies over the years which have demonstrated similar biologic activity. Inability to identify any species-specific transcripts further supports the similarity of the two populations.

-20-

The probe derived from the 3' end of baboon RNA recognized human cDNAs fairly well under appropriate hybridization conditions. The concentration of Cot1 and oligo-dT which are used for blocking non-specific hybridization were found to be very crucial for this purpose. This is not unexpected, because the genomes of the two species are highly conserved, and both have Alu sequences (Hamdi *et al.*, 2000; Hamdi *et al.*, 1999). In general, higher background resulting from the baboon probe may be a reflection that the Alu content is not identical, and might benefit from a readjustment of the hybridization conditions, especially Cot1 and oligo-dT concentration. Nonetheless, the hybridization signal obtained with the baboon probe was strong and resulted in a very similar pattern to the one obtained with human probe. This suggests that human cDNA arrays are accurate substrates for baboon experiments, thereby facilitating translation of experimental results with this animal model to human relevance.

The studies were performed using a cDNA filter array system and radioactive probes. Although there may be limitations to the use of filters rather than solid cDNA supports, GeneFilters were especially attractive for these studies because they contain over 25,000 different cDNA clones, which covers an estimated 50% of the human genome, including a large proportion of uncharacterized cDNAs (ESTs).

The use of GeneFilters dictated an experimental design that differs from those using cDNA arrays on solid supports. Because two probes cannot be simultaneously hybridized and compared in a single experiment, reproducibility is maximized when the same membrane is re-used for sequential hybridization to compare probes from different RNA sources. Due to limited membrane lifetime, it is not possible to repeat multiple experiments, or compare expression patterns among different subjects, so the sampling error may be greater than for other methods for cDNA analysis. Thus, the results presented here should be regarded as a starting point for further confirmation and analysis.

The most reliable data obtained on these filters is the comparison of relative signal strength for a single gene between two probes. An absolute determination of the relative expression between different genes on one filter is less reliable, because the signal strength is dependent on many factors, such as the length of the clone and the hybridization efficiency of the probe, and the relative inaccuracies of spotting

-21-

small amounts of DNA. Cross-comparisons of cDNA on different filters is less reliable. Here, the intensity of the hybridization signal relative to background was used as a means of comparison between filters, in order to estimate the relative level of expression of all of the genes on this dataset, recognizing that this is only an approximate-though generally reliable-measurement.

The gene list resulting from this study represents a selection of some of the most highly-abundant genes in hematopoietic cells, and provides a starting point to develop a profile of the predominant cDNAs that define CD34<sup>+</sup> cells. Interestingly, a significant fraction of the genes identified on these filters are not unique to hematopoietic cells, but are present in other tissues. This reinforces the concept that a tissue is defined not only by the expression of tissue-specific genes, but also by the overall pattern and relative abundance of the sequences which are more widely expressed. Perhaps the most interesting result is the fact that many of the cDNAs expressed at high level in these cells have not yet been identified or characterized. The gene and EST list presented here, and their relative expression levels, represent a potential wealth of new information about bone marrow stem cells and hematopoietic progenitor cells.

A comprehensive description of the CD34<sup>+</sup> transcriptome with reference to the UniGenes represented in GeneFilters will be useful. Although by no means complete, the list of over 15,000 cDNAs disclosed comprises an estimated 25 - 50% of the genes expressed in CD34<sup>+</sup> cells, and also provides an approximation of their relative abundance. This gene set will be useful for the production of customized cDNA arrays for bone marrow studies.

08697798.070201

-22-

## DOCUMENTS CITED

- Alizadeh *et al.* (2000) "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *"Nature"* 403:503-511.
- Andrews R-G, Singer J-W, Bernstein I-D. Monoclonal antibody 12-8  
5 recognizes a 115-kd molecule present on both unipotent and multipotent hematopoietic colony-forming cells and their precursors. *Blood.* 1986; 67:842-845.
- Andrews RG, Bryant EM, Bartelmez SH, et al. CD34<sup>+</sup> marrow cells, devoid of T and B lymphocytes, reconstitute stable lymphopoiesis and myelopoiesis in lethally irradiated allogeneic baboons. *Blood.* 1992;80:1693-1701.
- 10 Brandt JE, Galy AH, Luens KM et al. Bone marrow repopulation by human marrow stem cells after long-term expansion culture on a porcine endothelial cell line. *Exp. Hematol.* 1998; 26(10):950-61.
- Brandt JE, Bartholomew AM, Fortman JD, et al. Ex vivo expansion of autologous bone marrow CD34<sup>+</sup> cells with porcine microvascular endothelial cells  
15 results in a graft capable of rescuing lethally irradiated baboons. *Blood.* 1999;94:106-113.
- Goodell MA, Rosenzweig M, Kim H, et al. Dye efflux studies suggest that hematopoietic stem cells expressing low or undetectable levels of CD34 antigen exist in multiple species. *Nat. Med.* 1997;3:1337-1345.
- 20 Hamdi H, Nishio H, Zielinski R, Dugaiczky A. Origin and phylogenetic distribution of Alu DNA repeats: irreversible events in the evolution of primates. *J. Mol. Biol.* 1999;289: 861-871.
- Hamdi H-K, Nishio H, Tavis J, Zielinski R, Dugaiczky A. Alu-mediated phylogenetic novelties in gene regulation and development. *J. Mol. Biol.*  
25 2000;299: 931-939.
- Liao D, Pavelitz T, Weiner A-M. Characterization of a novel class of interspersed LTR elements in primate genomes: structure, genomic distribution, and evolution. *J. Mol. Evol.* 1998; 46: 649-660.
- Link H, Arseniev L, Bahre O, Kadar J-G, Diedrich H, Poliwoda H.  
30 Transplantation of allogeneic CD34<sup>+</sup> blood cells. *Blood.* 1996;87:4903-4909.
- Nachtman RG, Abdullah JM, Jurecic R. Cloning and functional characterization of novel genes preferentially expressed in hematopoietic cells

09897798-070201

-23-

[Abstract]. 29th Annual Meeting of the International Society for Experimental Hematology, Tampa, Florida:2000;28:108.

Phillips RL, Ernst RE, Brunk B, et al. The genetic program of hematopoietic stem cells. Science. 2000;288:1635-1640.

- 5        Pierelli L, Scambia G, Bonanno G, et al. CD34+/CD105+ cells are enriched in primitive circulating progenitors residing in the G0 phase of the cell cycle and contain all bone marrow and cord blood CD34+/CD38<sup>low/-</sup> precursors. Br. J. Haematol. 2000;108:610-620.

- 10        Trezise A-E, Godfrey E-A, Holmes R-S, Beacham I-R. Cloning and sequencing of cDNA encoding baboon liver alcohol dehydrogenase: evidence for a common ancestral lineage with the human alcohol dehydrogenase b-subunit and for class I *ADH* gene duplications predating primate radiation. Proc. Natl. Acad. Sci., U. S. A. 1989;86: 5454-5458 .

- 15        Ueda T, Yoshino H, Kobayashi K, et al. Hematopoietic repopulating ability of cord blood CD34<sup>+</sup> cells in NOD/Shi-scid mice. Stem Cells. 2000;18:204-213.

09897798.070201